Gene Expression Profiles in Normal and Cancer Cells

8 authors, including:

ette, IN) PM-80 pump. Separation was achieved by a BAS microbore column (MF-8949; 1 × 100 mm, with C18 packing of 3-μm particle size), which was attached directly to the injector (Rheodyne 9125) and to the UV detector (Waters 486 UV detector, outfitted with a Waters microbore cell kit). Adenosine was detected at a wavelength of 258 nm. Chromatographic data were recorded on a chart recorder, and the peak heights of microdialysis samples were compared to the peak heights of adenosine standards (1 pmol/10 μl) for quantification. The detection limit of the assay was 50 fmol (based on a signal-to-noise ratio of 3:1). Repeated assays of standards and pooled samples showed less than 10% variability. Custom-made CMA 10 probes from CMA/Microdialysis had a polycarbonate membrane (20,000-dalton cutoff), a 500-μm outer diameter, a 2-mm microdialysis membrane length, and a 35-mm shaft length. During the experiment, ACSF (composed of 147 mM NaCl, 3 mM KCl, 1.2 mM CaCl$_2$, and 1.0 mM MgCl$_2$, at a pH of 6.6) was pumped through the probe at a flow rate of 1.5 μl/min, the same flow rate used for drug perfusion. Consecutive 10-min dialysis samples were collected throughout the day via tubing with a low dead space volume (1.2 μl per 10 cm, FEP tubing; CMA/Microdialysis) and correlated with electrographically defined sleep-wakefulness states. Adenosine from a microdialysis sample produced a sharp chromatogram peak with a high signal-to-noise ratio and the same 8-min retention time as the adenosine standard (Fig. 1A).

13. For the analysis of the group data, a sleep cycle was defined as a continuous period that contained all of the behavioral states (W, SWS, and REM sleep), and began and ended with waking periods; the validity of comparisons over time was ensured by rejection of any cycles where there were suggestions of nonstationarity (adenosine values with >25% change between the first and last waking epochs). Of the samples in this comparison of W and SWS, 65% were 100% in a single state, and the remaining 35% had less than 20% of another state. The mean cycle duration was not different in the basal forebrain and thalamus samples.

14. NBTI actions are discussed in G. Sanderson and C. N. Scholfield [*Pfluegers Arch. Eur. J. Physiol.* **406**, 25 (1996)] and H. L. Haas and R. W. Greene [*Naunyn-Schmiedeberg's Arch. Pharmacol.* **337**, 561 (1988)]. These references and our preliminary data confirmed 1 μM as the lowest dose producing maximal effect. To ensure the presence of normal sleep, the 3-hour baseline period was not started until 30 min after the first REM episode (typically 1 to 2 hours after the animal was connected to the polygraph and microdialysis lines). Basal extracellular concentrations of adenosine were determined during the 3-hour baseline period that preceded the drug administration.

15. EEG power spectral analysis was performed during ACSF perfusion, during perfusion with 1 μM NBTI in the basal forebrain and thalamus, and during recovery sleep after 6 hours of wakefulness. Parietal EEG screw electrodes were used for EEG acquisition. The data were filtered at 70 Hz (low-pass filter) and 0.3 Hz (high-pass filter) with a Grass electroencephalograph and were continuously sampled at 128 Hz by a Pentium microprocessor computer with a Data-Wave (Data-Wave Technology, Longmont, CO) system. Absolute total power was calculated for the frequency range between 0.3 and 55 Hz. Five different frequency bands were used to calculate the relative power: delta, 0.3 to 4 Hz; theta, 4.1 to 9 Hz; alpha, 9.1 to 15 Hz; beta, 15.1 to 25 Hz; and gamma, 25.1 to 55 Hz. After basal forebrain NBTI perfusion, the relative power was significantly increased in the delta and decreased in the theta, alpha, beta, and gamma bands ($P < 0.04$; nonparametric Wilcoxon matched pairs signed-ranks test, used because of nonnormality of data). There was no change in power in any frequency band after NBTI infusion in the thalamus.

16. In evaluating the physiological relevance of adenosine at various concentrations, it is important to note that in vitro data from our laboratory (3) demonstrated that endogenous adenosine had a consistent inhibitory effect on cholinergic neurons. These data imply that adenosine's physiological effects in vivo are to be expected at baseline that is, without sleep deprivation or NBTI.

Rainnie et al. (3) did not measure endogenous adenosine concentrations, and thus the precise in vitro effects of doubling adenosine concentrations have not yet been specified, although it is known that there are progressive increases in inhibition of cholinergic neurons (beyond that seen from the endogenous inhibitory effect) with increasing concentrations of exogenously applied adenosine. Furthermore, we believe that the actions of adenosine that we have found in animal studies apply also to humans. First, the increase in EEG sleepiness with increasing duration of wakefulness has been documented in humans (1). Second, the adenosine physiology and pharmacology of experimental animals and of humans appear to be comparable [see reviews in (4–7) and also L. J. Findley, M. Boykin, T. Fallon, L. Belardinelli, *J. Appl. Physiol.* **65**, 556 (1988); and H. L. Haas, R. G. Greene, M. G. Yasargil, V. Chan-Palay, *Neurosci. Abstr.* **13**, 155 (1987)]. Finally, the adenosine antagonist caffeine increases wakefulness in formal experimental studies [see (7) and H. P. Landolt, D. J. Dijk, S. E. Gaus, A. A. Borbely, *Neuropsychopharmacology* **12**, 229 (1995)] and, as with the adenosine antagonist theophylline, constitutes the sleep-delaying ingredient in coffee and tea.

17. Changes in the entire relative power spectrum with NBTI infusion and in recovery sleep after prolonged wakefulness were, for each band, in the same direction (n = four animals).

18. P. H. Wu, R. A. Barraco, J. W. Phillis, *Gen. Pharmacol.* **15**, 251 (1984); R. Padua, J. D. Geiger, S. Dambock, J. I. Nagy, *J. Neurochem* **54**, 1169 (1990); J. G. Gu and J. D. Geiger, *ibid.* **58**, 1699 (1992). Both *N*-methyl-D-aspartate receptor agonists [C. G. Craig and T. D. White, *J. Pharmacol. Exp. Ther.* **260**, 1278 (1992); *J. Neurochem.* **60**, 1073 (1993)] and agonists that increase adenosine 3′,5′-monophosphate [R. W. Gereau and P. J. Conn, *Neuron* **12**, 1121 (1994); P. A. Rosenberg, R. Knowles, Y. Li, *J. Neurosci.* **14**, 2953 (1994)] might also increase extracellular adenosine concentrations by increasing extracellular adenine nucleotides that are ca-

tabolized to adenosine by 5′-ectonucleotidase (also a potential modulatory target).

19. This possibility has recently been reviewed by J. M. Brundege and T. V. Dunwiddie [*J. Neurosci.* **16**, 5603 (1996)], who also provided direct evidence for the possibility that an increase in intracellular adenosine (either by exogenous adenosine or inhibiting metabolism of endogenous adenosine) could lead to an increase in extracellular adenosine and its actions on receptors.

20. V. C. de Sánchez *et al., Brain Res.* **612**, 115 (1993); J. P. Huston *et al.*, *Neuroscience* **73**, 99 (1996). Adenosine appears to have a tighter linkage to sleep after wakefulness than do other putative SWS factors [see review by J. M. Krueger and J. Fang, in *Sleep and Sleep Disorders: From Molecule to Behavior*, O. Hayaishi and S. Inoue, Eds. (Academic Press and Harcourt Brace, Tokyo, Japan, in press)].

21. It is also possible that adenosine's effects in the neocortex may be directly attenuated by cholinergic receptor activation, as has been shown in the hippocampus [P. F. Worley, J. M. Baraban, M. McCarren, S. H. Snyder, B. E. Alger, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3467 (1987)]. Thus, adenosine's direct inhibitory effects on cholinergic somata might be enhanced by a consequent disinhibition of adenosine's effects on neocortical neurons. The specificity of sleep-wakefulness effects of NBTI does not support the idea that adenosine's effects result from a global action on brain neurons, as suggested by J. H. Benington and H. C. Heller [*Prog. Neurobiol.* **45**, 347 (1995)].

22. We thank P. Shiromani, D. Rainnie, and D. Stenberg for their advice during this work; L. Camara and M. Gray for technical assistance; and C. Portas for her preliminary work on this project. Supported by National Institute of Mental Health, grant R37 MH39, 683 and awards from the Department of Veterans Affairs to R.W.M.

3 March 1997; accepted 15 April 1997
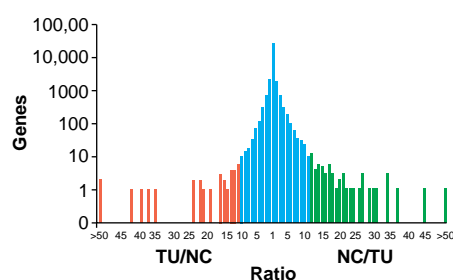
Gene Expression Profiles in Normal and Cancer Cells

Lin Zhang,* Wei Zhou,* Victor E. Velculescu, Scott E. Kern, Ralph H. Hruban, Stanley R. Hamilton, Bert Vogelstein, Kenneth W. Kinzler†

As a step toward understanding the complex differences between normal and cancer cells in humans, gene expression patterns were examined in gastrointestinal tumors. More than 300,000 transcripts derived from at least 45,000 different genes were analyzed. Although extensive similarity was noted between the expression profiles, more than 500 transcripts that were expressed at significantly different levels in normal and neoplastic cells were identified. These data provide insight into the extent of expression differences underlying malignancy and reveal genes that may prove useful as diagnostic or prognostic markers.

Much of cancer research over the past 50 years has been devoted to analyses of genes that are expressed differently in tumor cells as compared with their normal counterparts. Although hundreds of studies have pointed out differences in the expression of one or a few genes, no comprehensive study of gene expression in cancer cells has been reported. It is therefore not known how many genes are expressed differentially in tumor versus normal cells, whether the bulk of these differences are cell-autonomous rather than dependent on the tumor microenvironment, and whether most differences are cell type–specific or tumor-specific. Technological advances have made it possible to answer such questions through simultaneous analysis of the expression patterns of thousands of genes (1, 2). In this study, using normal and neoplastic gastrointestinal tissue as a prototype, we analyzed global profiles of gene expression in human cancer cells.

We used the recently developed method

Fig. 1. Comparison of expression patterns in CR cancers and normal colon epithelium. A semilogarithmic plot reveals 51 tags that were decreased more than 10-fold in primary CR cancer cells (green), whereas 32 tags were increased more than 10-fold (red); 62,168 and 60,878 tags derived from normal colon epithelium and primary CR cancers, respectively, were used for this analysis. The relative expression of each transcript was determined by dividing the number of tags observed in tumor and normal tissue as indicated. To avoid division by 0, we used a tag value of 1 for any tag that was not detectable in one of the samples. We then rounded these ratios to the nearest integer; their distribution is plotted on the abscissa. The number of genes displaying each ratio is plotted on the ordinate. TU, CR tumors; NC, normal colon.

Table 1. Overall summary of SAGE analysis.

	Normal colon	Colon tumors	Colon cell lines	Pancreatic tumors	Pancreatic cell lines	Total
Total tags	62,168	60,878	60,373	61,592	58,695	303,706
Genes*	14,721	19,690	17,092	20,471	14,247	48,741
GenBank†	8,753 (59)	10,490 (53)	10,193 (60)	11,547 (56)	8,922 (63)	26,339 (54)

*Indicates the number of different genes represented by the total tags analyzed (4). †Indicates the number of genes that match an entry in GenBank. Numbers in parentheses indicate the percentage of the total number of different tags.

Table 2. Summary of SAGE analysis by abundance classes.

Copies/cell	Normal colon	Colon tumors	Colon cell lines	Pancreatic tumors	Pancreatic cell lines	Total
>500						
Genes*	62 (29)	54 (25)	54 (19)	32 (11)	70 (26)	55 (19)
GenBank†	59 (95)	52 (96)	53 (98)	32 (100)	70 (100)	54 (98)
>50 and ≤500						
Genes*	645 (28)	470 (21)	618 (27)	657 (29)	585 (27)	59 (26)
GenBank†	545 (84)	429 (91)	579 (94)	609 (93)	529 (90)	553 (93)
>5 and ≤50						
Genes*	4,569 (27)	5,011 (29)	5,733 (34)	6,146 (36)	4,845 (31)	6,209 (30)
GenBank†	2,893 (63)	3,204 (64)	3,682 (64)	4,054 (66)	3,168 (65)	4,241 (68)
≤5						
Genes*	9,445 (16)	14,155 (25)	10,687 (20)	13,636 (24)	8,697 (16)	41,882 (25)
GenBank†	5,256 (56)	6,805 (48)	5,879 (55)	6,852 (50)	5,155 (59)	21,491 (51)

*For genes, the first number denotes the number of different genes (4) represented in the indicated abundance class. Numbers in parentheses indicate the mass fraction ($\times 100$) of total transcripts represented by the indicated abundance class. †For GenBank entries, the first number indicates the number of different genes that matched an entry in GenBank in the indicated abundance class. Numbers in parentheses indicate the corresponding percentage of genes.

called serial analysis of gene expression (SAGE) (2) to identify and quantify a total of 303,706 transcripts derived from human colorectal (CR) epithelium, CR cancers, or pancreatic cancers (Table 1) (3). These transcripts represented about 49,000 different genes (4) that ranged in average expression from 1 copy per cell to as many as 5300 copies per cell (5). The number of different transcripts observed in each cell population varied from 14,247 to 20,471. The bulk of the mRNA mass (75%) consisted of transcripts expressed at more than five copies per cell (Table 2). In contrast, most transcripts (86%) were expressed at less than five copies per cell, but in aggregate this low-abundance class represented only 25% of the mRNA mass. This distribution was consistently observed among the different samples analyzed and was consistent with previous studies of RNA abundance classes based on RNA-DNA reassociation kinetics ($R_o t$ curves) (6). Monte Carlo simulations revealed that our analyses

had a 92% probability of detecting a transcript expressed at an average of three copies per cell (7).

Many of the SAGE tags appeared to represent previously undescribed transcripts, as only 54% of the tags matched GenBank entries (Tables 1 and 2). Twenty percent of these matching transcripts corresponded to characterized mRNA sequence entries, whereas 80% matched uncharacterized expressed sequence tag (EST) entries. As expected, the likelihood of a tag being present in the databases was related to abundance; GenBank matches were identified for 98% of the transcripts expressed at >500 copies per cell but for only 51% of the transcripts expressed at ≤5 copies per cell. Because the SAGE data provide a quantitative assay of transcript abundance, unaffected by differences in cloning or polymerase chain reaction efficiency, these data provide an independent and relatively unbiased estimate of the current completeness of publicly available EST databases.

Comparison of expression patterns between normal colon epithelium and primary colon cancers revealed that most transcripts were expressed at similar levels (Fig. 1). However, the expression profiles also revealed 289 transcripts that were expressed at significantly different levels [$P < 0.01$ (8)]; 181 of these 289 were decreased in colon tumors as compared with normal colon tissue (average decrease, 10-fold; exam-

ples in Fig. 2A). Conversely, 108 transcripts were expressed at higher levels in the colon cancers than in normal colon tissue (average increase, 13-fold; examples in Fig. 2A). Monte Carlo simulations indicated that the analysis would have detected >95% of transcripts expressed at a sixfold or greater level in normal versus tumor cells or vice versa (9). Because relatively stringent criteria were used for defining differences [$P < 0.01$ (8)], the number of differences reported above is likely to be an underestimate.

To determine how many of the 289 differences were independent of the cellular microenvironment of cancers in vivo, we compared SAGE data from CR cancer cell lines with that from primary CR cancer tissues (10). Perhaps surprisingly, 130 of 181 transcripts that were expressed at reduced levels in cancer cells in vivo were also expressed at significantly lower levels in the cell lines (Table 3). Likewise, a significant fraction (47 of 108) of the transcripts expressed at increased levels in primary cancers were also expressed at higher levels in the CR cancer cell lines (Table 4). Thus, many of the gene expression differences that distinguish normal from tumor cells in vivo persist during in vitro growth. However, despite these similarities, there were also many differences. For example, only 47 of 228 genes expressed at higher levels in CR cancer cell lines were also expressed at high levels in the primary CR cancers.

L. Zhang, W. Zhou, B. Vogelstein, Howard Hughes Medical Institute, The Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA.
V. E. Velculescu, Oncology Center and Program in Human Genetics and Molecular Biology, The Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA.
S. E. Kern, R. H. Hruban, S. R. Hamilton, Department of Pathology and Oncology Center, The Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA.
K. W. Kinzler, Oncology Center, The Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed.

In combination, comparison of the expression pattern of CR cancer cells (in vivo or in vitro) to that of normal colon cells revealed 548 differentially expressed transcripts (Tables 3 and 4). The average difference in expression for these transcripts was 15-fold. Although the ability to detect differences is influenced by the magnitude of the variance, with the power to detect smaller differences being less, 92 transcripts that were less than threefold different were identified among the 548 transcripts. However, the genes exhibiting the greatest differences in expression are likely to be the most biologically important.
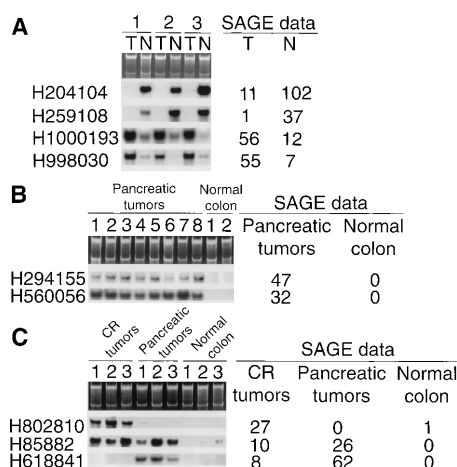
To determine whether the changes noted in CR cancers were neoplasia- or cell type–specific, we performed SAGE on mRNA derived from pancreatic cancers. A total of 404 transcripts were expressed at higher levels in pancreatic cancers as compared with normal colon epithelium (examples in Fig. 2B). Most (268) of these transcripts were pancreas-specific (*11*) (see example in Fig. 2C), although 136 were also expressed at high levels in CR cancers. These 136 transcripts constituted 47% of the 289 transcripts that were increased in CR cancers relative to normal colon tissue and are likely to be related to the neoplastic process rather than to the specific cell type of origin.

One question that arose from these data is the potential heterogeneity of expression between individual tumors. The SAGE data were acquired from two samples of each tissue type (normal colon, primary CR cancer, CR cancer cell line, and so on). To examine the generality of these expression profiles, we arbitrarily selected 27 differentially expressed transcripts and evaluated them in 6 to 12 samples of normal colon and primary cancers by Northern (RNA) blot analysis (*12*). In general, expression patterns were very reproducible among different samples. Of 10 genes with elevated expression in normal colon relative to CR cancers as determined by SAGE, each was detected in the normal colon samples and was expressed at considerably lower levels in tumors (Fig. 2A). Similarly, most of the genes identified by SAGE as increased in CR or pancreatic cancers were confirmed to be reproducibly expressed in most primary cancers examined by Northern blot analysis (Fig. 2, A and B). It is important to note, however, that there were differences among the cancers, with a few cancers exhibiting particularly large or small amounts of individual transcripts. Such differences in gene expression undoubtedly contribute to the observed heterogeneity in the biological properties of cancers derived from the same organ (*13*).

What are the identities of the differentially expressed genes? Of the 548 differentially expressed transcripts, 337 were tentatively identified through database comparisons. When tested, most (93%) of these identifications proved to be legitimate (*14*), as was expected from previous SAGE analyses (*2*). Although a large number of differentially expressed genes were identified, some simple patterns did emerge. For example, genes that were expressed at higher levels in normal colon

Fig. 2. Northern blot analysis of genes differentially expressed in gastrointestinal neoplasia. Northern blot analysis was performed on total RNA (5 μg) isolated from primary CR carcinomas (T) and matching normal colon epithelium (N) or pancreatic carcinomas. The top line of gels in each panel shows ethidium bromide–stained gels before transfer. The number of SAGE tags observed in the original analysis is indicated to the right of each blot. (**A**) Examples of transcripts that were decreased or increased in CR cancers. (**B**) Examples of transcripts increased in pancreatic cancers (*11*). (**C**) Examples of transcripts increased in cancer that were or were not cancer type–specific. The following probes were used for Northern blot analysis [human SAGE tag identifier, gene product name (GenBank accession number)]: (A) H204104, guanylin (M95714); H259108 (see Tables 3 and 4); H1000193 (see Tables 3 and 4); H998030 (see Tables 3 and 4). (B) H294155, RIG-E (U42376); H560056, TIMP-1 (S68252). (C) H802810, EST338411 (W52120); H85882, 1-8D (X57351); H618841, GA733-1 (X13425). An additional 19 examples of Northern blots are available on the Internet at http://welchlink.welch.jhu.edu/~molgen-g/home.htm.

Table 3. Transcripts decreased in CR cancer. The 20 transcripts displaying the largest decrease in expression in CR cancers (in vivo and in vitro) are listed by fold reduction. The tag sequence represents the 10–base pair SAGE tag, and SAGE UID is the human SAGE tag identifier. Probable GenBank matches are listed and those in boldface were confirmed by Northern blot analysis or by cloning and sequence analysis. Fold changes in expression were calculated as described in Fig. 1. TU, colon tumors; CL, colon cell lines; NC, normal colon. Tables of all 548 differentially expressed genes are available on the Internet at http://welchlink.welch.jhu.edu/~molgen- g/home.htm.

Tag sequence	SAGE UID	NC/TU	TU	CL	NC	GenBank match (accession number)
GACCAGTGGC	H545514	45	1	0	45	No match
ATTTCAAGAT	H259108	37	1	0	37	**Carbonic anhydrase II (M36532)**
GTCATCACCA	H740629	34	0	0	34	Uroguanylin (U34279)
CTTATGGTCC	H511670	34	1	0	34	No match
TGGAAAGTGA	H950457	34	1	1	34	**Human cellular oncogene c-*fos* (V01512)**
CCTTCAAATC	H390158	31	1	0	31	**Carbonic anhydrase I (M33987)**
TCGGAGCTGT	H893564	30	1	4	30	EST 261490 (H98618)
GTCTGGGGGA	H752297	29	1	3	29	**EST 81394 (T60135)**
GATCCCAACT	H578824	27	1	1	27	Metallothionein from cadmium-treated cells (V00594)
CTTAGAGGGG	H510123	27	1	5	27	No match
ATGATGGCAC	H233106	26	0	2	26	No match
CCTGTCTGCC	H388582	24	1	2	24	EST 122594 5′ (T99568)
CTGGCAAAGG	H500747	23	0	0	23	No match
CTTGACATAC	H516402	22	0	0	22	***Homo sapiens* CL100 mRNA for protein tyrosine phosphatase (X68277)**
GGAAGAGCAC	H657554	21	1	1	21	**Gal-β (1-3/1-4)GlcNAc α-2.3-sialyltransferase (X74570)**
TCTGAATTAT	H909556	21	1	1	21	Transmembrane carcinoembryonic antigen BGPb (X14831)
TAAATTGCAA	H790417	19	6	1	113	Cytokeratin 20 (X73502)
GTGGGGGCGC	H764570	18	1	1	18	EST 153570 5′ (R48529)
ATGGTGGGGG	H241323	18	2	6	36	*Homo sapiens* zinc finger transcriptional regulator mRNA (M92843)
TCACCGGTCA	H857781	17	7	7	122	Human mRNA for plasma gelsolin (X04412)

epithelium than in CR tumors were often related to differentiation. These genes included fatty acid–binding protein (*15*), cytokeratin 20 (*16*), carbonic anhydrase (*17*), guanylin (*18*), and uroguanylin (*19*), which are known to be important for the normal physiology or architecture of colon epithelium (Tables 3 and 4). On the other hand, genes that were increased in CR cancers were often related to the robust growth characteristics that these cells exhibit. For example, gene products associated with protein synthesis, including 48 ribosomal proteins, five elongation factors, and five genes involved in glycolysis were observed to be elevated in both CR and pancreatic cancers as compared with normal colon cells. Although most of the transcripts could not have been predicted to be differentially expressed in cancers, several have previously been shown to be dysregulated in neoplastic cells. The latter included IGFII (*20*), B23 nucleophosmin (*21*), the Pi form of glutathione-S-transferase (*22*), and several ribosomal proteins

(*23*), all of which were increased in cancer cells, as previously reported. Likewise, Dra (*24*) and gelsolin (*25*) were decreased in cancer cells, as previously reported. Surprisingly, two widely studied oncogenes, c-*fos* and c-*erbb*3, were expressed at much higher levels in normal colon epithelium than in CR cancers, in contrast to their up-regulation in transformed cells (*26*).

These data provide basic information necessary for understanding the gene expression differences that underlie cancer phenotypes. They also provide a necessary framework for interpreting the significance of individual differentially expressed genes. Although this study demonstrated that a large number of such differences exist (about 500 at the depth of analysis used), it was equally remarkable that the fraction of transcripts exhibiting significant differences was relatively small, representing 1.5% of the transcripts detected in any given cell type (*27*). The fact that many, but not all, of the differences were preserved during in vitro culture demonstrates the utility of cul-

tured lines for examination of some aspects of gene expression but also provides a note of caution about relying on such lines to perfectly mimic tumors in their natural environment. Finally, the finding that hundreds of specific genes are expressed at different levels in CR cancers, and that some of these are also expressed differentially in pancreatic cancers, provides a wealth of reagents for future biologic and diagnostic experimentation.

Table 4. Transcripts increased in CR cancer. The 20 transcripts displaying the greatest increase in CR cancers (in vivo and in vitro) are listed by fold induction. Conditions are as described in Table 3.

Tag sequence	SAGE UID	TU/NC	TU	CL	NC	GenBank match (accession number)
CTTGGGTTTT	H518912	73	73	42	0	**Insulin-like growth factor II splice form 1 (IGFII) (X07868)**
TACAAAATCG	H802871	42	42	20	0	**Insulin-like growth factor II splice form 2 (IGFII) (X07868)**
GTGTGTTTGT	H769020	24	24	15	0	**TGF-β-induced gene *Beta-igh3* (M77349)**
AAAAGAAACT	H2056	16	16	27	1	**Human mRNA for poly(A) binding protein (Y00345)**
TGCTGCCTGT	H948604	15	15	16	1	*H. sapiens* HCG IV mRNA (X81005)
CTGATGGCAG	H495251	14	14	15	0	EST 324128 3′ (W46476)
GCCCAAGGAC	H610466	12	12	19	0	**Human mRNA for actin-binding protein (filamin) (X53416)**
ACTCGCTCTG	H121311	12	12	16	0	EST 342926 3′ (W67797)
ATCTTGTTAC	H229106	11	11	28	0	Human mRNA for fibronectin (FN precursor) (X02761)
AAGCTGCTGG	H40571	10	10	17	0	Isoform 1 gene for L-type calcium channel, exons 41 and 41A (Z26305)
TGAAATAAAA	H918273	9	18	37	2	Human *hB23* gene for B23 nucleophosmin (X16934)
TTATGGGATC	H998030	8	55	78	7	**Human MHC protein homologous to chicken B complex (M24194)**
CAATAAATGT	H274492	7	60	73	9	**Human mRNA for ribosomal protein L37 (D23661)**
CTCCTCACCT	H482584	6	72	41	12	Human *Bak* mRNA, complete cds (U16811)
ACTGGGTCTA	H125661	6	29	25	5	*H. sapiens* RNA for *nm23-H2* gene (X58965)
CTGTTGATTG	H507455	5	44	54	9	**Human liver mRNA fragment DNA binding protein UPI (X04347)**
TTCAATAAAA	H1000193	5	56	154	12	**Human acidic ribosomal phosphoprotein P1 mRNA (M17886)**
AAGAAGATAG	H33331	4	39	69	9	Human ribosomal protein L23a mRNA, partial cds (U02032)
CTGGGTTAAT	H502724	4	115	160	29	*H. sapiens* S19 ribosomal protein mRNA, completed (M81757)
CTGTTGGTGA	H507577	4	65	116	17	Human homolog of yeast ribosomal protein S28 (D14530)

REFERENCES AND NOTES

1. M. D. Adams *et al.*, *Nature* **377** (suppl. 28), 3 (1995); M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **270**, 467 (1995); J. Derisi *et al.*, *Nature Genet.* **14**, 457 (1996); T. M. Gress *et al.*, *Oncogene* **13**, 1819 (1996); D. J. Lockhart *et al.*, *Nature Biotechnol.* **14**, 1675 (1996); M. Schena *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10,614 (1996).
2. V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, *Science* **270**, 484 (1995); V. E. Velculescu *et al.*, *Cell* **88**, 243 (1997).
3. To minimize individual variation, approximately equal numbers of tags (30,000) were derived from two different patients for each tissue. For primary tumors (two CR carcinomas and two pancreatic adenocarcinomas), RNA was isolated from portions of tumors judged by histopathology to contain 60 to 90% tumor cells. The cells grown in vitro were derived from CR (SW837 and Caco2) and pancreatic (ASPC-1 and PL45) cancer cell lines. CR epithelial cells were isolated from sections of normal colon mucosa from two patients with the use of EDTA as described [S. Nakamura, I. Kino, S. Baba, *Gut* **34**, 1240 (1993)]. Histopathology confirmed that the isolated cells were >90% epithelial. Isolation of polyadenylate RNA and SAGE was performed as described (*2*). SAGE data were analyzed with SAGE software and GenBank Release 94 as described (*2*).
4. A total of 69,393 different SAGE tags were identified among the 303,706 tags analyzed. A small fraction of these different tags was likely due to sequencing errors. SAGE analysis of yeast (*2*), for which the entire genomic sequence is known, demonstrated a sequencing error rate of ~0.7%, translating to a SAGE tag error rate of 6.8% ($1 - 0.993^{10}$). Because these sequencing mistakes are essentially random, they do not substantially affect the analysis, although they could artificially inflate the number of different genes identified. Therefore, to be conservative, we reduced our estimate of different genes identified by this maximum tag error rate (that is, 6.8% of 303,706 total tags). The number of different tags derived from the same gene because of alternative splicing was assumed to be negligible.
5. Abundance can be determined simply by dividing the observed number of tags for a given transcript by the total number of tags obtained. An estimate of about 300,000 transcripts per cell was used to convert the abundances to copies per cell [N. D. Hastie and J. O. Bishop, *Cell* **9**, 761 (1976)].
6. J. O. Bishop, J. G. Morton, M. Rosbash, M. Richardson, *Nature* **250**, 199 (1974); B. Lewin, *Gene Expression* (Wiley, New York 1980), vol. 2.
7. Computer simulations indicated that analysis of 300,000 tags would yield a 92% chance of detecting a tag for a transcript whose expression on average was at least three copies per cell among the tissues examined, assuming 300,000 transcripts per cell.
8. To minimize the number of assumptions and to account for the large number of comparisons being made, we used Monte Carlo analysis to determine statistical significance. The null hypothesis was that the level, kind, and distribution of transcripts were the same for cancer and normal cells. For each transcript, we performed 100,000 simulations to determine the relative likelihood, due to chance alone (*p*-chance), of obtaining a difference in expression equal to or greater than the observed difference, given the

null hypothesis. We converted this likelihood to an absolute probability value by simulating 40 experiments in which a representative number of transcripts (27,993 transcripts in each experiment) were identified and compared. We derived the distribution of transcripts used for these simulations from the average level of expression observed in the original samples. We then compared the distribution of the p-chance scores obtained in the 40 simulated experiments (false positives) with those obtained experimentally. On the basis of this comparison, a maximum value of 0.0005 was chosen for p-chance. This yielded a false-positive rate that was no higher than 0.01 for the least significant p-chance value below the cutoff.

9. Two hundred simulations, assuming an abundance of 0.0001 in one sample and 0.0006 in a second sample, revealed a significant difference [$P < 0.01$ (8)] 95% of the time.

10. This analysis revealed 208 transcripts that were significantly decreased in CR cancer cell lines as compared with normal colon cells and 228 transcripts that were increased. Venn diagrams and tables illustrating the relation between the in vivo and in vitro differences are available through the Internet at http://welchlink.welch.jhu.edu/~molgen-g/home.htm.

11. It is not possible to obtain pancreatic duct epithelium, from which pancreatic carcinomas arise, in sufficient quantities to perform SAGE. It is therefore not possible to determine whether these transcripts were derived from genes that were highly expressed

only in pancreatic cancers or that were also expressed in pancreatic duct cells.

12. Total RNA isolation and Northern blot analysis were performed as described [W. S. el-Deiry et al., Cell **75**, 817 (1993)].

13. A. H. Owens, D. S. Coffey, S. B. Baylin, Eds., *Tumor Cell Heterogeneity: Origins and Implications* (Academic Press, New York, 1982).

14. Northern blot analyses were done on 45 of the 337 differentially expressed transcripts with tentative database matches. In three cases, the pattern of expression was not differentially expressed as predicted by SAGE and, for the purposes of this calculation, they were presumed to represent incorrect database matches.

15. D. C. Rubin, D. E. Ong, J. I. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 1278 (1989); K. Okubo, J. Yoshii, H. Yokouchi, M. Kameyama, K. Matsubara, *DNA Res.* **1**, 37 (1994).

16. R. Moll et al., *Differentiation* **53**, 75 (1993).

17. J. Sowden, S. Leigh, I. Talbot, J. Delhanty, Y. Edwards, *ibid.*, p. 67.

18. F. J. de Sauvage et al., *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9089 (1992).

19. R. C. Wiegand et al., *FEBS Lett.* **311**, 150 (1992).

20. J. V. Tricoli et al., *Cancer Res.* **46**, 6169 (1986); S. Lambert, J. Vivario, J. Boniver, R. Gol-Winkler, *Int. J. Cancer* **46**, 405 (1990).

21. W. Y. Chan et al., *Biochemistry* **28**, 1033 (1989).

22. J. D. Hayes and D. J. Pulford, *Crit. Rev. Biochem. Mol. Biol.* **30**, 445 (1995).

23. G. F. Barnard et al., *Cancer Res.* **52**, 3067 (1992); P. J. Chiao, D. M. Shin, P. G. Sacks, W. K. Hong, M. A. Tainsky, *Mol. Carcinogen* **5**, 219 (1992); N. Kondoh, C. W. Schweinfest, K. W. Henderson, T. S. Papas, *Cancer Res.* **52**, 791 (1992); G. F. Barnard et al., *ibid.* **53**, 4048 (1993); M. G. Denis et al., *Int. J. Cancer* **55**, 275 (1993); J. M. Frigerio et al., *Hum. Mol. Genet.* **4**, 37 (1995).

24. C. W. Schweinfest, K. W. Henderson, S. Suster, N. Kondoh, T. S. Papas, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4166 (1993).

25. M. Tanaka et al., *Cancer Res.* **55**, 3228 (1995); D. Medina, F. S. Kittrell, C. J. Oborn, M. Schwartz, *ibid.* **53**, 668 (1993).

26. A. D. Miller, T. Curran, I. M. Verma, *Cell* **36**, 51 (1984); M. H. Kraus, W. Issing, T. Miki, N. C. Popescu, S. A. Aaronson, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9193 (1989).

27. In the case of normal and neoplastic colon cancer tissue, 548 differentially expressed transcripts were identified among the 36,125 different transcripts.

28. We thank K. Polyak and P. J. Morin for providing colon cancer cell lines; G. M. Nadasdy for providing pancreatic primary tumors; and J. Floyd, C. R. Robinson, and Y. Beazer-Barclay for technical assistance. Supported by the Clayton Fund and by NIH grants GM07309, CA57345, and CA62924. B.V. is an investigator of the Howard Hughes Medical Institute.

21 January 1997; accepted 25 March 1997

Brain Regions Responsive to Novelty in the Absence of Awareness

Gregory S. Berns,* Jonathan D. Cohen, Mark A. Mintun

Brain regions responsive to novelty, without awareness, were mapped in humans by positron emission tomography. Participants performed a simple reaction-time task in which all stimuli were equally likely but, unknown to them, followed a complex sequence. Measures of behavioral performance indicated that participants learned the sequences even though they were unaware of the existence of any order. Once the participants were trained, a subtle and unperceived change in the nature of the sequence resulted in increased blood flow in a network comprising the left premotor area, left anterior cingulate, and right ventral striatum. Blood flow decreases were observed in the right dorsolateral prefrontal and parietal areas. The time course of these changes suggests that the ventral striatum is responsive to novel information, and the right prefrontal area is associated with the maintenance of contextual information, and both processes can occur without awareness.

The detection of novelty is a cognitive operation necessary to survival and requires an assessment of both expectedness and context. Events can be familiar in one context but novel in another. More precisely, novelty represents a deviation from the expected likelihood of an event on the basis of both previous information and internal estimates of conditional probabilities (1).

G. S. Berns, Department of Psychiatry, University of Pittsburgh Medical Center, Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, PA 15213, USA. J. D. Cohen, Department of Psychology, Carnegie Mellon University, and Department of Psychiatry, University of Pittsburgh Medical Center, Pittsburgh, PA 15213, USA. M. A. Mintun, Departments of Radiology and Psychiatry, Washington University School of Medicine, St. Louis, MO 63110, USA.

*To whom correspondence should be addressed.

Novelty detection has typically been linked to consciousness because novel events often capture attention. For similar reasons, studies of novelty have often been confounded by awareness (2). Here, we sought to determine whether the response to novelty can occur without awareness and, if so, to identify the associated brain regions in a manner unconfounded by awareness. To do so, we used an implicit learning task.

A large body of research has examined learning mechanisms that operate below the level of awareness. This type of learning is said to occur implicitly because behavioral measures indicate that learning takes place, even though the individuals are unaware of this or are unable to report it explicitly (3). A frequently used paradigm is

based on a serial reaction-time task, in which participants observe sequences of visual stimuli and must press buttons corresponding to these. Unknown to the participants, the sequence of stimuli is predetermined by a fixed, repeating order. With practice, reaction times improve (compared with randomly sequenced stimuli), indicating that the participants have learned about the sequential order. However, they are not always conscious of this. When the sequence is sufficiently complex, individuals are unaware of the sequential regularities or that they have learned anything specific about the stimuli, even though their reaction times have improved significantly (4). This indicates that sequential information can be both learned and used in the absence of awareness.

One type of sequence that has been well studied is based on finite-state grammars (5). Such grammars can be used to generate highly complex, context-dependent sequences. With enough practice, individuals show improvements consistent with implicit learning of such grammars. However, because such grammars are typically probabilistic, specific repeating sequences rarely occur, further reducing the likelihood of awareness of the sequential regularities.

Implicit learning of finite-state grammars means that participants have developed expectations for each stimulus, on the basis of the specific stimuli that preceded it in the sequence (that is, its context). Under such conditions, changing the rules of the grammar will cause subsequent stimuli to violate these expectations, by appearing in novel contexts. Thus, a switch in grammars